Examining the Relationship Between Alcohol and Mortality Through the Use of Propensity Scores

Anita L. Iyer

Masters Thesis, Department of Health Management and Health Economics

UNIVERSITETET I OSLO

December 2011

Examining the Relationship Between Alcohol and Mortality Through The Use Of Propensity Scores

2011

Examining the Relationship Between Alcohol and Mortality Through The Use Of Propensity Scores

Anita L. Iyer

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

Introduction: The relationship between alcohol and health has long been of interest to researchers. Over the past few decades, the risk associated with alcohol consumption has often been characterized as a J-shaped curve. Studies within this area have largely concluded that moderate alcohol consumption is associated with lower mortality. Despite the evidence in support of the J-curve hypothesis, the debates continue, suggesting that analytical approaches beyond standard survival analysis can be explored.

Objective: To re-examine the relationship between moderate alcohol consumption and all-cause mortality through the utilization of propensity score matching, and to examine the effect the passage of time and the algorithm used have on the outcome.

Results: When a matching algorithm which allowed for replacement was utilized, the difference in the risk of mortality for lifelong abstainers and moderate drinkers was no longer significantly different following a series of adjustments. By contrast, when replacement was restricted, the mortality risk for lifelong abstainers remained significantly greater ($p>0.05$), though it was reduced by a large margin. Sensitivity analyses revealed that if unobserved variables caused the odds ratios to differ by a factor of 1.25 (2009) or 1.15 (2003) the results would no longer be significant, suggesting that unobserved or excluded variables could reshape the results.

Conclusions: The results of this analysis generally support the findings of much of the published literature in this area. Six additional years of mortality data served to offer more support for the J-curve hypothesis; however, this additional support is only apparent when replacement is restricted. This analysis underscores the importance the matching algorithm used can play in shaping the results.

Acknowledgements

Writing a Masters Thesis can feel like a solitary endeavor, but the process does not occur in a vacuum. The faculty and staff at the Department and Health Management and Health Economics have been instrumental in instilling the skills and providing the support needed for completing this project. I'd like to thank my advisor, Hans Olav Melberg, who introduced me to this general area of analysis, and whose guidance has made this process educational and on occasion, entertaining. Thanks also to Birthe Neset, for all of her support over the past few years.

In the time we have spent studying together, my fellow students in the International Masters Program have been a constant source of inspiration and insight. Special thanks to David Mwaura, who pointed me in the direction of this topic, Søren Klitkou for his Stata wizardry, and Guri Lise Furnes Pettersen and Sanaz Arshi for their friendship, support, and company during the long hours we spent together in the computer room.

Finally, to my friends and family, who have been there for me in ways I never imagined – thank you. No names, but remembered, thanked, and loved, all the same.

Anita Lakshmi Iyer

December 2011

Table of Contents

List of Abbreviations and Acronyms

CHD	Coronary Heart Disease
HALS	Health And LifeStyle Survey - Complete Dataset (HALS1, HALS2; And Survival Data)
HALS1	Health And LifeStyle Survey
HALS2	Health And LifeStyle Survey Seven Years On
NHS	National Health Service
NN	Nearest Neighbor
PS	Propensity Score
PSM	Priopensity Score Matching
RGSC	Registrar General Social Class

List of Tables and Figures

1 Introduction

1.1 Alcohol Consumption and Health

Alcohol has long been used for its medicinal, antiseptic and recreational properties, so it is unsurprising that modern medical science has dedicated resources to the study of alcohol consumption's subsequent effects on health. One area of particular interest has been the relationship between alcohol and mortality in general, which has been studied for over a century. A historical overview reveals that the effect of alcohol consumption on various cardiovascular conditions and long-term mortality varies, both in regards to each individual condition and in terms of the amount of alcohol consumed (Klatsky 2002).

However, a number of studies have shown that moderate alcohol consumption result in lower all-cause mortality, particularly when contrasted with outcomes observed in both those who abstain from alcohol consumption and those who drink heavily (Ellison 2002, Corrao et al. 2000). In particular, the published advantages of moderate alcohol consumption encompass a range of benefits, including reduced rates of myocardial infarction and heart failure and reduced risk for ischemic stroke, dementia, and osteoporosis, among others. By contrast, heavy alcohol consumption negatively affects a wide array of physical and mental functions (Standridge, Zylstra & Adams 2004).

While the link between moderate alcohol consumption and lower CHD mortality has been documented, the exact mechanism at work remains unclear. A recent Norwegian population based cohort study found that, while alcohol intake is related to a reduced risk of coronary heart disease, the serum level of high density cholesterol is not part of the causal pathway associated with this relationship (Magnus et al 2011). This area continues to remain a topic of interest for clinicians and researchers alike.

1.2 The J-curve Hypothesis and the Implications for Public Health

The findings within this area of research have been used to argue that the relationship between alcohol and health can be visualized as a J-shaped curve. This curve exists when an inverse relationship between an established risk factor and a given health outcome is observed to a certain point, wherein the more commonly associated positive relationship between the risk factor and assumed outcome reasserts itself (Goetghebeur and Pocock 1995). A 2006 meta analysis, which pooled data from 34 prospective studies available in Pubmed as of December 2005 concluded that the consumption of up to 4 drinks per day in men and 2 drinks per day in women, was inversely associated with total mortality, with the maximum protection being 18% in women and 17% in men. By contrast, higher levels of consumption increased the mortality risk, leading the authors to conclude that a J-shaped relationship between alcohol and total mortality was confirmed in these adjusted studies (Di Castelnuovo et al 2006).

Naturally, the existence of a J-shaped curve for alcohol consumption raises particular concerns. Some researchers, embracing a J-curve hypothesis and its implications, have suggested health care providers counsel their patients accordingly. However, despite the published benefits, the debate over integrating such findings into treatment regimens or public health guidelines continues. In late 2010, a letter to the editor (Ponz de Leon 2010) entitled "What should we advise about alcohol consumption?" published in the journal *Internal and Emergency Medicine* sparked what a January 18, 2011 ScienceDaily article referred to as "a debate among scientists." The editorial, and its subsequent responses, highlight the difficult decisions faced by physicians who must weigh the ethics of recommending moderate alcohol consumption despite the potential risks of excessive use against the notion of withholding advice in an attempt to avoid the dangers of excess.

As the recent debates suggest, the relationship has real implications for those seeking to counsel patients. As noted previously, a substantial number of prospective studies conducted on alcohol use indicate that those who abstain completely from alcohol use carry a greater mortality risk than those who consume moderate amounts of alcohol. This additional mortality burden applies to both "all cause" mortality in general, and coronary heart disease in particular. However, Fillmore et al's (2006) meta analysis tested the extent to which individuals who had a previous history of alcohol consumption had been systematically and erroneously misclassified as abstainers, and whose present abstention could be associated with aging or ill-health. The studies which appeared to be free of the aforementioned classification errors did not demonstrate a significant degree of cardiac or all-cause protection attributable to moderate alcohol consumption, which suggests that the health benefits of moderate consumption found in previous studies has been overestimated. In the years following this analysis, some of the concerns raised by Fillmore et al. have been addressed. A study published earlier this year in *Demography*, which sought to specifically avoid these and other classification errors, found that individuals who consume moderate alcohol have both lower all-cause mortality and CHD mortality (Fuller 2011). However, the debate surrounding the issue continues.

1.3 Objectives of this Thesis

The primary objective of this thesis is to re-examine the relationship between moderate alcohol consumption and all-cause mortality through the utilization of propensity score matching, and to see whether the results of a propensity score analysis reflect the same support for the J-curve hypothesis as the regression and survival analysis techniques which are dominant in the published literature within this area. This thesis can be seen as an extension of the analysis originally published by Melberg in 2006, with the benefit of an additional 6 years' worth of mortality data. A secondary aim of this thesis is to examine how this additional time period and data influence the outcomes.

1.4 Organization of this Thesis

The first section of this thesis has framed the research question within a greater public health debate, and stated the objective. Section two will present the argument for the use of propensity score matching within this context. Section three will present the data used for this analysis, and present the arguments used for selecting the dataset, individual cases, and variables used in the analysis. Section four will present the results, and Section five will present a discussion of these results and conclusions.

2 Analytical Approach

2.1 Searching for Causality in Observational Data

2.1.1 Defining Causality

Studying the effects of long-term alcohol consumption patterns poses certain challenges when it comes to the issue of causation. A causal effect is defined as the difference in outcome Y for individual i attributable to a given treatment. Given outcome Y, where Y_0 denotes the outcome if not treated and Y_1 denotes outcome if treated, the causal effect for individual i is the difference between Y_1 and Y_0, or

$$\Delta(i) = Y_1(i) - Y_0(i) \quad (1)$$

In this particular case, the health outcome can be defined as the risk of dying before June 2009. In this situation, Y_1 is the risk associated with a lifetime pattern of moderate alcohol consumption, whereas Y_0 implies the risk following a lifetime of abstention. However, it is impossible to determine true causation within this framework, as individual i can only be observed in one state. An individual is either a moderate drinker or a lifelong abstainer; he or she cannot exist in both groups simultaneously. Thus, it is impossible to compare the health outcome for each of these treatments directly.

Since it is impossible to observe an individual in two states simultaneously, other approaches to determining causality have to be explored. One approach would be to base our conclusions on the observed average differences between the two groups. In this case, one could simply identify a group of moderate drinkers, compare them to a group of lifelong abstainers, and draw conclusions from their comparative health outcomes. However, alcohol consumption, like smoking, exercising, and a host of other health-related behavior, is influenced by a variety of personal factors. It is these factors which complicate the issue. In order for this approach to work, the two groups would have to be identical in every possible way save one: their approach to alcohol consumption. However, reality deviates from this ideal, and individuals who choose to abstain from alcohol have likely chosen this approach for a reason. For example, assume that lifelong abstainers are older and more likely to suffer from illness than moderate drinkers. Merely comparing the average mortality rate would be misleading, as illness and old age automatically increase an individual's likelihood of dying earlier, regardless of alcohol consumption

This example of selection bias explains the appeal and popularity of randomized experiments. Once a sample population is selected, the randomization of subjects to different treatments works to prevent systematic bias in the observed and unobserved covariates (D'Agustino 1998). However, unlike a new drug or diet program, there is no ethical or feasible way to design an experimental study which assigns individuals to a lifetime regime of alcohol consumption or abstinence. Under these circumstances, the utilization of observational data becomes key.

2.1.2 Expected Effect and Selection Bias

Statistical approaches to the measurement of causation in observational studies have been reviewed in several studies (Holland 1986; Rosenbaum 1999).

When utilizing observational data, an emphasis is placed on the counterfactual framework. which assumes that individuals observed in a predetermined control or treatment state have potential outcomes in the corresponding counterfactual state, i.e. that, despite only being observed in one state, individuals have potential outcomes in all (Winship & Morgan 1999). However, there is always a missing piece of the puzzle.

One approach to circumvent the problem posed by a reality where individuals can only be observed in one particular state is to restrict the focus to the overall expected effect of a treatment on the population, instead of on a given individual. For example, one could consider the overall expected effect of a lifetime of moderate alcohol consumption:

$$E(\Delta) = E(Y_1\ ) - E(Y_0\ ) \quad (2)$$

However, this premise merely measures the difference in mortality rates among moderate drinkers and lifelong abstainers; it does not actually show what would happen if the individuals in each group had somehow been assigned to a lifelong drinking pattern. Nor does it address the fundamental problem many raise with using observational data, which is sample selection bias. Sample selection bias may arise in practice for two reasons: self selection by the individuals included in the sample or sample selection decisions made by data analysts (Heckman 1979). When considering the effect of long term alcohol consumption on mortality, one must assume that individuals, to some degree, have selected a "treatment" for themselves – i.e., they have chosen to live as a drinker or a lifelong abstainer.

Let us imagine that *D* indicates that an individual receives a treatment, where *D*=1 is moderate drinking and *D*=0 implies lifelong abstention. Let X represent some individual characteristics which are relevant for the outcome, such as age, gender, or race (Todd 2006). The outcomes observed can be expressed as

$$E[Y_1|D=1, X]$$

$$(3)$$

$$E[Y_0|D=0, X]$$

where the first expression indicates the outcome for the moderate drinkers while the second expression is the outcome for the lifelong abstainers. Again, there is an unobserved component, as to compensate for selection bias, one is also interested in the outcome for the treated if they had not been treated, and vice versa. These unobservable outcomes can be expressed as:

$$E[Y_1|D=0, X]$$

$$(4)$$

$$E[Y_0|D=1, X]$$

Heckman et al (1998, p.2) characterized the selection bias that arises through the use of a comparison group which utilizes data on "nonparticipants" to form an estimate. In this context, using this method assumes that, conditional on X, the outcomes of lifelong abstainers approximate what moderate drinkers would have experienced had they also chosen to abstain; that is, it assumes

$$E[Y_0|D=0, X] \cong E[Y_0|D=1, X] \quad (5)$$

However, if this assumption fails to hold, the selection bias associated with factor X is

$$E[Y_0|D=1, X]- E[Y_0|D=0, X] \quad (6)$$

In the case of a lifetime pattern of alcohol consumption, it is impossible to assume that the conditions in (5) are met. In essence, the causal effect of moderate alcohol consumption

$$E[Y_1|D=1, X]- E[Y_0|D=1, X] \quad (7)$$

is difficult to separate from the selection bias expressed in (6).

2.2 Counteracting Selection Bias

2.2.1 Regression vs. Matching

Statistical methods which attempt to deal with the selection bias in observational studies can be broadly categorized as belonging to two groups (Melberg 2006, p.197). The first of these two categories consists of regression techniques. Regression analyses generally seek to find a set of control variables that can be included in a regression equation in order to remove the correlation between the treatment variable and the error term (Winship and Morgan, 1999). While some regression methods account for unobserved covariates, these methods are often limited by assumptions regarding the functional form and distribution of data, and can risk extreme extrapolation (Melberg 2006, p.198).

The second group of methods is based on the idea of matching. The essential logic behind a matching approach is intuitively appealing; comparing "like with like" is an approach which is easy to understand. Matching methods do not rely on the same assumptions of linearity and error distributions as regression methods. However, the reliance on sub groups presents another problem. Often referred to as the "2^n Problem" or the curse of dimensionality, the issue is simple: as the number of binary variables one wishes to analyze increases, the number of sub groups needed increases exponentially. A model which only adjusts for one variable, (for example, gender) would only require two sub groups. However, to examine gender, the presence or absence of a university degree, and marital status would demand 2^3, or 8 groups. As more variables are added, the number of groups becomes so large that each group lacks the requisite number of observations to draw reliable conclusions

2.2.2 Rosenbaum & Rubin’s Propensity Score

Propensity score matching presents a solution to the 2^n problem. The intuitive appeal of a matching mechanism is that it mimics a randomized experiment by matching individuals on the characteristics - more formally, the covariates - which are assumed to impact the outcome. In Rosenbaum and Ruben's seminal 1983 article, a propensity score is defined as the "coarsest function" of the covariates which is a balancing score, b(x), where b(x) "is a function of the observed covariates x such that the conditional distribution of x" given b(x) is the same for treated (D = 1) and control (D = 0) units (p.42).

Adapting the notation used by Rosenbaum and Ruben, the propensity score for individual *i* is

$$e(x_i) = pr(D_i = 1|X_i = x_i) \quad (8)$$

where it is assumed that for the given X's, the D's are independent.

$$pr(D_1 = d_1, \dots, D = D_N|X_1 = x_1, \dots, X_N = x_N) = \prod_{i=1}^{N} e(x_i)^{z_i}\{1 - e(x_i)\}^{1-d_1} \quad (9)$$

In essence, a propensity score creates a single variable using the relevant covariates X. This variable, denoted P, is the propensity score. Formally, an individual's propensity score is defined as "the probability of being treated conditional on (or based only on) the individual's covariate values" (D'Agustino 1998, p. 2266). More simply, the propensity score is the probability an individual receives treatment D, given X. In this context, a propensity score gives us the probability that given X, where X includes factors such as age, gender, social integration, etc. which we assume impacts an individual's alcohol consumption pattern and mortality. By comparing individuals with similar propensity scores, one can estimate what the outcome for a given individual would have been if they had chosen a different lifestyle. Essentially, this exercise provides the missing pieces of information from equation (4), and mimics the mechanism used in a randomized experiment.

2.2.3 Nearest Neighbor Matching

Implementing propensity score matching requires the selection of a matching algorithm. Nearest neighbor (NN) matching has been described as the most "straightforward" matching approach (Caliendo and Kopeinig, 2008, p.41).The concept behind NN matching is simple: each individual from the treatment group is matched with the individual in the control group with the closest propensity score. Variations on NN matching are largely contingent on making a trade off concerning bias or variance (ibid), This trade off is illustrated when weighing the relative merits of one-to-one matching with or without replacement. When matching is done without replacement, it means that each individual in the control group can only be matched with one individual from the treatment group. In this case, it means that each moderate drinker can only serve as a match for one lifelong abstainer. Under these conditions, the variance decreases, but the bias may increase, particularly if there is a reason to suspect that the propensity scores among the two groups are distributed in such a way that individuals in the treatment group will have a limited number of individuals in the control group with similar propensity scores. By contrast, allowing for replacement means that the same individual in the control group can be used as a match by more than one individual in the treatment group. While this reduces bias by providing the possibility for the best possible match for each individual, it has consequences for the variance.

3 Data, Case, and Variable Selection

3.1 The Health and Lifestyle Survey (HALS)

The dataset selected for analysis is the Health and Lifestyle Survey (HALS), an extensive study conducted by the University of Cambridge Clinical School in the UK. The data set includes three parts. The original data set (HALS1), which consists of a representative sample of 9,003 individuals, was assembled using interviews, physiological measurements taken by a nurse, and a self-report booklet during 1984/85 (Cox 1988). A follow up study, HALS2, was conducted in 1991/92. At this time, 5,352 individuals, or 59.4% of the original participants, participated in this follow-up study (Cox 1995). A third data file, containing mortality data on the individuals from the HALS dataset includes data generated through June 2009. This mortality data was generated by tagging individuals in the British National Health Service (NHS) and updated automatically as participants' health outcomes are entered into the system (Cox 2009).

3.2 Moderate Drinkers and Lifelong Abstainers: Criteria for Categorization and Case Selection

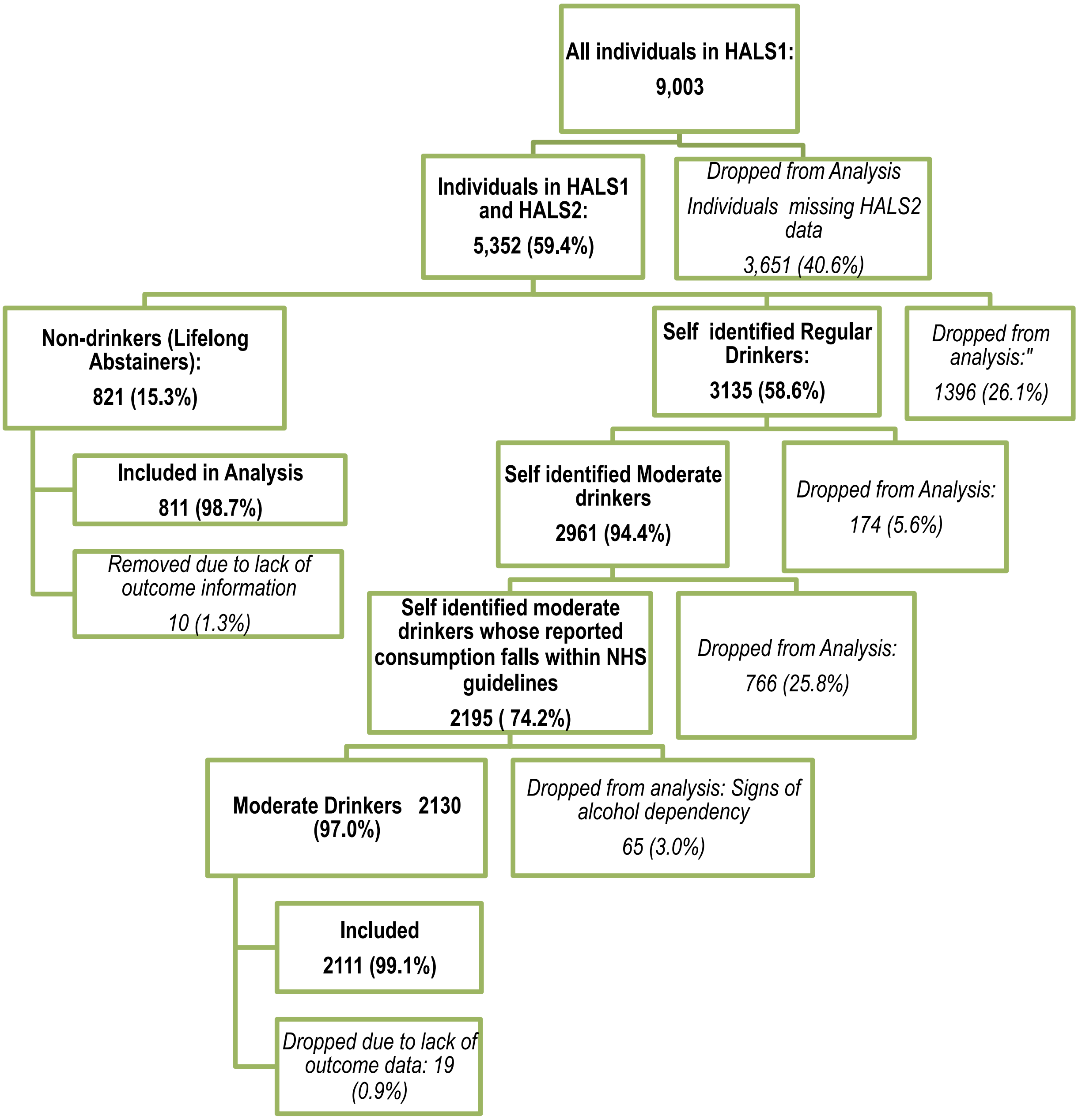
3.2.1 Criteria for Case Selection

Longitudinal studies have shown that for a majority of individuals, patterns of alcohol consumption vary throughout a lifetime, making patterns of consumption difficult to measure over a long term (Temple & Fillmore 1985). While the HALS dataset does not account for a lifetime's worth of measurements, only individuals who participated in both the HALS 1 and HALS2 studies were selected for analysis. Research on the reliability and validity of self report measures as a measure of alcohol consumption have generally concluded that such measures are generally reliable, provided the data has been collected in a way that minimizes bias (Midanik 1988, Del Boca & Darkes 2003). Certain criteria were set for variables related to alcohol consumption, and the process of selection was designed to filter out those individuals whose responses changed over time or lacked internal consistency. The statistical package PASW was used to select cases which fulfilled these criteria. For a full overview of the variables used to select cases for analysis, please see Table 1.

3.2.2 Variable Selection and Categorization

The breadth of the HALS dataset made variable selection a priority. As the majority of responses are based on an individual's recollection, in many cases, attempts to elicit the same information in different ways have been used. For this analysis, individuals who do not consume alcohol and those who are considered long-term moderate drinkers were included. This meant that individuals who used to drink (ex drinkers), or those whose alcohol use could not be described as moderate (heavy drinkers) were removed from the analysis. Individuals whose drinking habits also changed, such as individuals who self-reported as non-drinkers at HALS1 but who reported as moderate drinkers in the second study, were also removed from the analysis. In addition, a small percentage of individuals have no mortality data available and had to be excluded. In total, 2,922 individuals, or 54.6% of those who participated in the HALS1 & HALS2 surveys, were selected for analysis. Of these, 811 (27.8%) individuals have been identified as lifelong abstainers, while 2,111(72.2 %) are considered moderate drinkers.

Figure 1: Process of Case Selection



“Dropped from Analysis” indicates that individual cases were excluded due to not meeting selection criteria. These groups include individuals who were interviewed at HALS1 but not at HALS2, former drinkers, individuals whose drinking habits shifted between HALS1 and 2, self identified heavy drinkers, self-identified moderate consumers whose consumption exceeded NHS limits for moderate consumption, individuals whose self-reported behavior suggested an unhealthy dependence on alcohol, and individuals who were not categorized due to missing data.

3.2.3 Identification of Lifelong Abstainers

The HALS questionnaire included a series of questions designed to pinpoint alcohol consumption habits. One of these questions was a self-report measure, where individuals were asked to describe themselves as non-drinkers, special occasion drinkers, occasional drinkers, or regular drinkers. Those identifying as non-drinkers or special occasion drinkers were then asked if they had always been abstainers. Individuals who identified as lifelong abstainers or special occasions drinkers in both HALS1 and HALS2 were considered non-drinkers. Several of the questions used to categorize drinkers were used as a secondary control, to ensure that these individuals had been identified correctly.

3.2.4 Identification of Moderate Drinkers

One of the biggest challenges posed by the HALS dataset is the subjective nature of many of the responses. As previously noted, the HALS questionnaire included a self-report measure, where individuals described themselves as non-drinkers, special occasion drinkers, occasional drinkers, or regular drinkers. A second question asked both current and former occasional and regular drinkers whether they categorized their drinking habits as light, moderate or heavy, or whether they didn't know. However, one individual's regular drinking may be considered heavy drinking by the health authorities, while another individual's regular drinking may be limited to a few beers each month with colleagues, which may perceived as special occasion drinking by others.

The HALS questionnaires also requested participants who self-identified as drinkers to fill out a "drink diary", which documented the amount of wine, beer, and spirits each individual had consumed in the last week. Individuals were also asked whether their past week's drinking behavior reflected their normal consumption habits. While this information is also dependent on an individual's recall and honesty, it is possible to compare the units of alcohol consumed in a week against published public health guidelines. As the HALS data was collected within the UK, guidelines from the British National Health System served as a comparison point. These guidelines stress that men should not exceed 3-4 units of alcohol per day, or 28 units of alcohol per week, with a unit defined as 10ml of pure alcohol (NHS 2011). Women are encouraged to consume no more than 2-3 units of alcohol per day, or no more than 21 units per week.

Unfortunately, the HALS alcohol units were measured in a less precise fashion, and one can argue that the units used in HALS are not equivalent to the NHS units. Nonetheless, in this analysis, individuals consuming more than the recommended NHS weekly totals have been excluded from the moderate drinkers category.

Thus, individuals categorized as life-long moderate drinkers are those whose responses reflect moderate alcohol consumption patterns, both in their categorization of themselves and in terms of the amounts of alcohol they report consuming in the past week, and on a regular basis.

3.2.5 Exclusion Criteria

The individuals excluded completely from the analysis fall into three categories. As this analysis aims to compare lifelong abstainers with lifelong moderate drinkers, individuals who have adopted patterns of abstention prior to participation in the survey or between HALS1and HALS2 have been excluded. Likewise, individuals whose responses show signs of unhealthy consumption – either in terms of the amount of alcohol consumed per week, or other problematic behavior, have also been excluded. Finally, individuals missing data on their drinking habits or missing outcome data have been excluded.

Table 1: Variables used to isolate Lifelong Abstainers and Moderate Drinkers			
Variable name	**Corresponding HALS2 Variable**	**Description**	**Comments**
ALSELF	ALSELF2	Self-identified alcohol consumption pattern	
ALEX	ALEX2	Identification of those who have quit drinking as opposed to lifelong abstainers	Individuals who identified as ex-drinkers were not considered abstainers
ALEXAGE	ALEXAGE2	Age individual stopped drinking (If applicable)	Positive responses were used to exclude individuals from the pool selected for analysis, as positive responses indicate changed drinking patterns
ALSELFQ	ALSELFQ2	Description of current/former drinking habits	
ALQ101	*ALQBEER1*	*Number of alcohol units per week - Beer*	*These three variables were combined to create AlUnitTotal. a measure of total consumption over a week's time*
ALQ102	*ALQWINE1*	*Number of Alcohol Units per week – Wine*	
ALQ103	*ALQSPIR1*	*Number of alcohol units per week – Spirits*	
ALPROB3	alpr203	Having a drink first thing in the morning as nerve/hangover cure	This question was used to help filter out those who may demonstrate signs of dependency as drinking first thing in the morning has been highlighted as a key indicator of a drinking problem (Israel et. al 1996)
AlUnitTotal	AlUnitTotal2	Derived from ALQ101, ALQ102, ALQ103, ALQBEER1, ALQWINE1, the total Total Units per week	NOTE. Units from HALS roughly correspond to NHS guidelines, but are not equivalent. Thus, this variable has been used as a secondary control in conjunction with ALSELF and ALSELFQ
ALTYPIC		Does past week's consumption reflect normal consumption patterns?	

3.3 Identifying Relevant Covariates

Having identified lifelong moderate drinkers and lifelong abstainers within the sample, the next stage of the analysis centered on selecting relevant covariates. Some factors, such as age and gender, have been so well established as being of general interest within public health and medical research that they are automatically included. One of the arguments often raised against the use of propensity score matching (and other methods which rely on the statistical analysis of observational data) is the largely subjective nature of covariate selection. While it is impossible to avoid subjective decision-making, the literature does provide support for the selection of certain factors. An overview of the covariates, and the variables utilized in their construction, is provided in the Appendix.

3.3.1 Self Reported Global Health Measure

One important component of the HALS dataset is that it includes several self-reported measures of health. Self ratings of health status encompass and assess a large number of covariates, and can provide insight which goes beyond the scope of more specific covariate measures. Idler and Benyamini's (1997) review of 27 community studies found that global self-rated health acted as an independent predictor of mortality, regardless of the inclusion of numerous specific health status indicators and other relevant covariates known to predict mortality. More specifically. self reported or subjective measures of health have been found to be associated with mortality due to diabetes, infectious and respiratory diseases, heart disease, stroke, and cancer (Benjamins et. al 2004).

3.3.2 Self Reported Risk Factors

While the HALS dataset also includes certain standard measures compiled by nurses, such information is not available for all the participants. Of the cases selected for analysis, 340 (11.6%) were not visited and evaluated by a nurse when the HALS data were collected. By contrast, the vast majority of individuals provided information about their medical history. In addition to questions about an individual's overall health status, more targeted questions about the presence of particular risk factors, such as smoking, suffering from a chronic illness, presence or absence of a handicap, and a history of conditions such as high blood pressure or heart disease have been included.

3.3.3 Socioeconomic Status

Attempts to incorporate measures of socioeconomic status into an analytic framework has lead to a variety of approaches, due to the array of factors which play a role in shaping an individuals socioeconomic status. Factors such as income, education, household assets, and employment can all play a role. Historically, countries have differed in how they measure an individual's socioeconomic status. Feinstein's (1993) literature review underscored that an individual's occupation has often served as the measure of socioeconomic status in the United Kingdom, and that studies have revealed that health outcomes have historically differed amongst the different social classes. Social class measures within the HALS dataset are based largely on the Registrar General Social Class classification, which divided the population into six major groups based on occupation type.

3.3.4 Measures of Social Integration

Social interaction is an intrinsic part of our lives, and its influence on health outcomes is well documented. A meta analytic review of 148 studies concluded that the influence of social relationships on risk for mortality is comparable with other well-established risk factors for mortality (Holt-Lunstad, Smith & Layton 2010). Furthermore, the authors concluded that the association was strongest for more complex measures of social isolation, and weakest for simple binary measures. For example, a question such as "Do you live alone?" overlooks the fact that an individual living alone may have a supportive social network not captured by this simple binary response. Likewise, an individual living with a roommate can experience intense feelings of isolation, despite appearing better connected. This analysis incorporates responses to questions identifying feelings of loneliness, isolation, and a lack of social support, as well as looking at the more concrete measures of the size of a given individual's household.

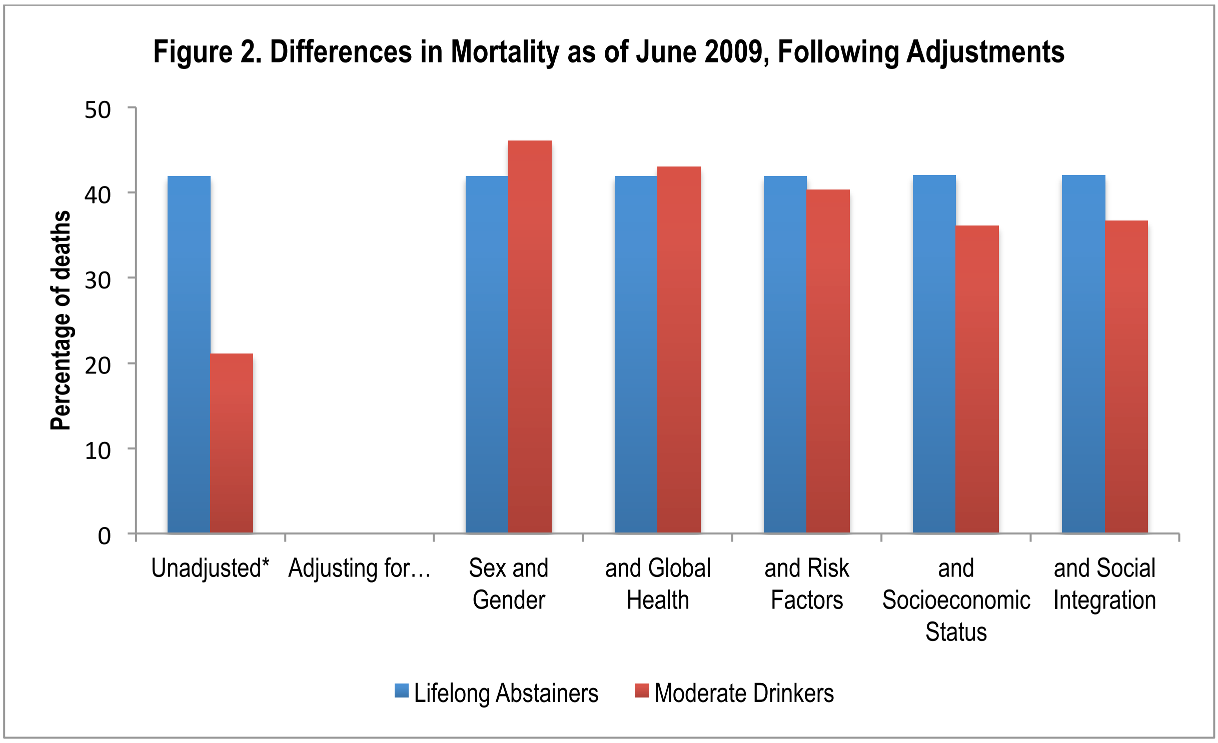
4 Results

4.1 The Simple Approach – Mortality Among Lifelong Drinkers and Abstainers

Of the 2,922 individuals included in the analysis, 786 (26.8%) had died as of June 2009. Simply dividing the sample into lifelong abstainers and moderate drinkers reveals that mortality appears to be much higher among the lifelong abstainers included in the sample. 78. 8% of moderate drinkers were still alive in June 2009, nearly 25 years after they had first participated in the HALS survey. By contrast, only 58.1% of lifelong abstainers included had survived until that point.

However, this relatively simplistic view does not consider any of the many confounders which could have contributed to this difference, the most obvious factors including the age and gender of the sample. Moderate drinkers had a mean age of 43.1 in 1984/1985, while the average age among lifelong abstainers was 53.1 - a full decade higher. With this in mind, the fact that a higher percentage of lifelong abstainers had passed away approximately 25 years later is not particularly surprising.

4.2 Adjustment Using One-to-One Matching With Replacement



** Denotes result where difference was significant.*

Prior to adjustment, abstainers are 20.8% more likely to be dead than moderate drinkers. The first stage of the analysis adjusted for two of the most obvious parameters of interest, mainly age and gender. The applied work was done in Stata, and utilized the psmatch2 command developed by Leuven and Sianesi in 2003. Following this adjustment, lifelong abstainers are 4.2% less likely to be dead compared to their matched counterparts. However, as discussed earlier, there are many other factors which the literature suggests contribute to alcohol consumption patterns and mortality.

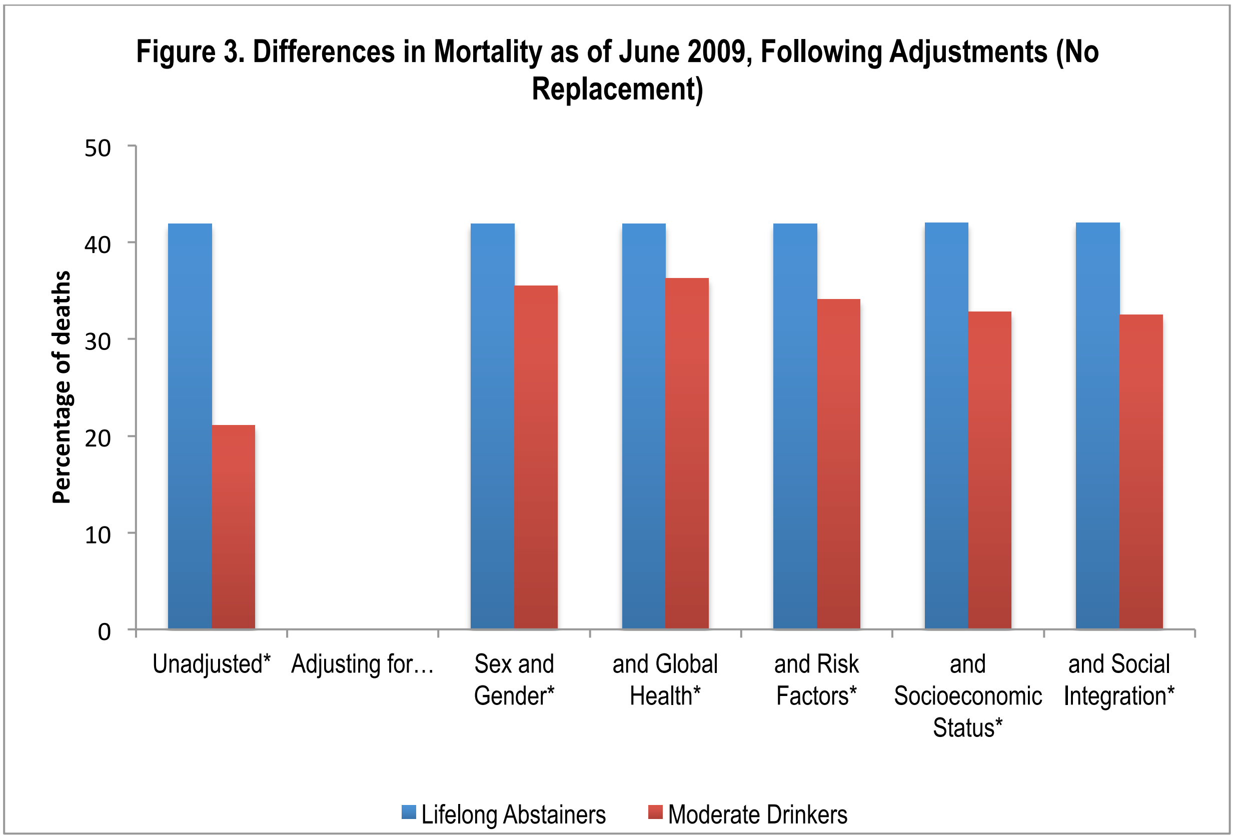
The next stage of the analysis incorporated a global health measure, which resulted in lifelong abstainers being 1.1% less likely to be dead in 2009. Following adjustments for self-reported risk factors, socioeconomic status, and social integration, lifelong abstainers were 5.3% more likely to be dead than their moderate drinking counterparts. However, this difference is no longer considered significant. In essence, when comparing similar individuals, the benefits associated with moderate consumption are much less apparent. The matching process also served to reduce the bias observed within the two groups. Table 2 provides an overview of the variables included in this analysis. For the vast majority of variables, bias has been strongly reduced or eliminated following the matching procedure.

Table 2: Overview of Important Confounding Variables, Averages, and Bias Reduction (HALS1, HALS2, and 2009 Mortality Data)							
Analysis 1: NN Matching, with replacement							
	Variable	**Sample**	**Mean: Lifelong Abstainers**	**Mean: Moderate Drinkers**	**Difference (%)**	**Reduction in Bias (%)**	**p value**
General	Age	Unmatched	53.077	43.020	64.3		0.000
		Matched	53.077	52.767	2.0	96.9	0.699
	Gender	Unmatched	0.200	0.336	-31.3		0.000
		Matched	0.200	0.229	-6.8	78.3	0.146
Global Health	Poor Health	Unmatched	0.431	0.316	23.9		0.000
		Matched	0.431	0.456	-5.2	78.4	0.317
Risk Factors	Low Activity Level	Unmatched	0.092	0.071	7.4		0.065
		Matched	0.092	0.071	7.7	-3.6	0.121
	Considers Self Overweight	Unmatched	0.166	0.120	13.2		0.001
		Matched	0.166	0.169	-0.7	94.6	0.894
	Prescription Drug Use - 1984/85	Unmatched	0.431	0.292	29.4		0.000
		Matched	0.431	0.435	-0.8	97.3	0.880
	Prescription Drug Use 91/92	Unmatched	0.564	0.423	28.5		0.000
		Matched	0.564	0.556	1.5	94.7	0.764
	Restricted Diet-1984/1985	Unmatched	0.102	0.072	10.4		0.009
		Matched	0.102	0.093	3.1	70.4	0.557
	Restricted Diet-1991/92	Unmatched	0.134	0.096	12.0		0.003
		Matched	0.134	0.160	-8.2	31.9	0.140
	Handicapped	Unmatched	0.318	0.202	26.7		0.000
		Matched	0.318	0.332	-3.1	88.3	0.559
	Suffers from Chronic Illness	Unmatched	0.483	0.399	17.0		0.000
		Matched	0.483	0.493	-2.0	88.2	0.691
	Reported High Blood Pressure	Unmatched	0.291	0.207	19.4		0.000
		Matched	0.291	0.254	8.6	55.6	0.094
	Reported Heart Trouble	Unmatched	0.104	0.046	22.1		0.000
		Matched	0.104	0.121	-6.6	70.1	0.271
	Smokes/Has Smoked	Unmatched	0.493	0.661	-34.4		0.000
		Matched	0.493	0.473	4.1	88.2	0.426
	Treated for Depresson	Unmatched	0.207	0.199	1.9		0.647
		Matched	0.207	0.223	-4.0	-111.8	0.431
	Left School Before 16	Unmatched	0.694	0.506	39.0		0.000
		Matched	0.694	0.685	1.8	95.4	0.707
	Lower Status	Unmatched	0.204	0.104	28.0		0.000
		Matched	0.204	0.202	0.7	97.5	0.902

Table 2 (continued)							
	Variable	**Sample**	**Mean: Lifelong Abstainers**	**Mean: Moderate Drinkers**	**Difference (%)**	**Reduction in Bias (%)**	**p value**
Social Integration	Feelings of Loneliness	Unmatched	0.331	0.354	-4.9		0.242
		Matched	0.331	0.341	-2.1	57.1	0.674
	Felt Unloved	Unmatched	0.031	0.014	11.6		0.002
		Matched	0.031	0.032	-0.8	92.8	0.887
	No Friends Nearby	Unmatched	0.102	0.086	5.5		0.178
		Matched	0.102	0.093	3.0	45.7	0.557
	Felt Isolated 1991/92	Unmatched	0.139	0.162	-6.4		0.126
		Matched	0.139	0.140	-0.3	94.6	0.943
	Felt Isolated (84/85)	Unmatched	0.181	0.189	-2.2		0.602
		Matched	0.181	0.187	-1.6	26.5	0.748
	Household Size - 1984/85	Unmatched	1.915	2.170	-17.0		0.000
		Matched	1.915	1.975	-4.0	76.3	0.424
	Household Size- 1991/92	Unmatched	1.491	1.811	-23.7		0.000
		Matched	1.491	1.589	-7.3	69.4	0.140

4.3 Adjustments using NN matching without replacement

4.3.1 Reduction in Bias



** Denotes adjustments where the resulting difference was significant.*

The second part of the analysis sought to explore the impact of modifying the matching algorithm on the results. In contrast to the first analysis, the difference in the percentage of likely deaths remained significant between the two groups despite the inclusion of various covariates of interest. Following the same process of covariate adjustment, lifelong abstainers remained 9.8% more likely to have died than their drinking counterparts. While this difference is a sizable reduction from the 20.8% difference which existed before adjustment, this difference is still significant ($p<0.05$) and implies that abstention from alcohol serves to increase an individual's risk of dying before June 2009.

Table 3 presents a summary of the covariates. Covariates which remain significantly different ($p<.0.05$) following the matching procedure are bolded.

Table 3: Overview of Important Confounding Variables, Averages, and Bias Reduction (HALS1, HALS2, and 2009 Mortality Data)							
Analysis 2: NN matching, without replacement							
	Variable	**Sample**	**Mean: Lifelong Abstainers**	**Mean: Moderate Drinkers**	**Difference (%)**	**Reduction in Bias (%)**	**p value**
General	**Age**	**Unmatched**	**53.077**	**43.020**	**64.3**		**0.000**
		Matched	**53.077**	**50.592**	**15.9**	**75.3**	**0.001**
	Gender	Unmatched	0.200	0.336	-31.3		0.000
		Matched	0.200	0.230	-7.1	77.4	0.130
Global Health	Poor Health	Unmatched	0.431	0.316	23.9		0.000
		Matched	0.431	0.405	5.4	77.3	0.290
Risk Factors	Low Activity Level	Unmatched	0.092	0.071	7.4		0.065
		Matched	0.092	0.081	4.1	45.2	0.425
	Considers Self Overweight	Unmatched	0.166	0.120	13.2		0.001
		Matched	0.166	0.159	2.1	83.9	0.686
	Prescription Drug Use - 1984/85	**Unmatched**	**0.431**	**0.292**	**29.4**		**0.000**
		Matched	**0.431**	**0.382**	**10.4**	**64.5**	**0.043**
	Prescription Drug Use 91/92	Unmatched	0.564	0.423	28.5		0.000
		Matched	0.564	0.527	7.5	73.6	0.134
	Restricted Diet-1984/1985	Unmatched	0.102	0.072	10.4		0.009
		Matched	0.102	0.084	6.2	40.8	0.230
	Restricted Diet-1991/92	Unmatched	0.134	0.096	12.0		0.003
		Matched	0.134	0.125	2.7	77.3	0.604
	Handicapped	Unmatched	0.318	0.202	26.7		0.000
		Matched	0.318	0.290	6.5	75.5	0.214
	Suffers from Chronic Illness	Unmatched	0.483	0.399	17.0		0.000
		Matched	0.483	0.455	5.8	66.1	0.252
	Reported High Blood Pressure	Unmatched	0.291	0.207	19.4		0.000
		Matched	0.291	0.269	5.2	73.4	0.319
	Reported Heart Trouble	Unmatched	0.104	0.046	22.1		0.000
		Matched	0.104	0.090	5.2	76.5	0.356
	Smokes/Has Smoked	Unmatched	0.493	0.661	-34.4		0.000
		Matched	0.493	0.522	-5.9	83	0.252
	Treated for Depression	Unmatched	0.207	0.199	1.9		0.647
		Matched	0.207	0.214	-1.8	2.2	0.714
	Left School Before 16	Unmatched	0.694	0.506	39.0		0.000
		Matched	0.694	0.673	4.4	88.8	0.363
	Lower Status	**Unmatched**	**0.204**	**0.104**	**28.0**		**0.000**
		Matched	**0.204**	**0.162**	**11.8**	**58**	**0.029**

Table 3: (continued)							
	Variable	**Sample**	**Mean: Lifelong Abstainers**	**Mean: Moderate Drinkers**	**Difference (%)**	**Reduction in Bias (%)**	**p value**
Social Integration	Feelings of Loneliness	Unmatched	0.331	0.354	-4.9		0.242
		Matched	0.331	0.323	1.6	67.8	0.750
	Felt Unloved	Unmatched	0.031	0.014	11.6		0.002
		Matched	0.031	0.025	4.2	63.9	0.450
	No Friends Nearby	Unmatched	0.102	0.086	5.5		0.178
		Matched	0.102	0.089	4.3	22.4	0.397
	Felt Isolated 1991/92	Unmatched	0.139	0.162	-6.4		0.126
		Matched	0.139	0.154	-4.2	35.3	0.398
	Felt Isolated (84/85)	Unmatched	0.181	0.189	-2.2		0.602
		Matched	0.181	0.185	-1.0	55.9	0.847
	Household Size - 1984/85	**Unmatched**	**1.915**	**2.170**	**-17.0**		**0.000**
		Matched	**1.915**	**2.069**	**-10.3**	**39.5**	**0.042**
	Household Size- 1991/92	**Unmatched**	**1.491**	**1.811**	**-23.7**		**0.000**
		Matched	**1.491**	**1.656**	**-12.2**	**48.6**	**0.015**

4.3.2 Sensitivity Analysis

This result differs from Melberg's (2006) findings, and from the findings explored in the preceding section. Naturally, this raises several questions. While the HALS dataset provides a great deal of information, it is impossible to take every possible covariate of interest into account. Secondly, it appears apparent that in this sample, the two groups differ significantly in terms of the distribution of several key covariates. It is entirely feasible that the results have been shaped by factors of interest which have been unobserved, unmeasured or simply excluded. A sensitivity analysis was performed through the utilization of the mhbounds command in Stata, which checks the sensitivity of the estimated average treatment effects on the treated by computing Mantel-Haenszel bounds (Becker and Caliendo 2007). Due to an assumption of independence, this analysis is only appropriate when matching without replacement has been used, as matching with replacement will lead to biased results. Results of the analysis are provided in Table 4. Bolded values indicate significance ($p<0.05$).

Table 4: Mantel-Haenszel (1959) bounds for The Risk of Dying Before June 2009

Gamma (Γ)	Q_mh+	Q_mh-	p_mh+	p_mh-
1	3.91191	3.91191	**0.000046**	**0.000046**
1.05	3.43966	4.38728	**0.000291**	**5.70E-06**
1.1	2.9888	4.8401	**0.0014**	**6.50E-07**
1.15	2.55843	5.27353	**0.005257**	**6.70E-08**
1.2	2.14674	5.68923	**0.015907**	**6.40E-09**
1.25	1.75213	6.08868	**0.039875**	**5.70E-10**
1.3	1.37321	6.47318	0.084843	**4.80E-11**
1.35	1.00874	6.84388	0.156549	**3.90E-12**
1.4	0.657634	7.20178	0.255387	**3.00E-13**
1.45	0.318909	7.54781	0.374898	**2.20E-14**
1.5	-0.008303	7.88277	0.503312	**1.60E-15**

Gamma : odds of differential assignment due to unobserved factors
Q_mh+ : Mantel-Haenszel statistic (assumption: overestimation of treatment effect)
Q_mh- : Mantel-Haenszel statistic (assumption: underestimation of treatment effect)
p_mh+ : significance level (assumption: overestimation of treatment effect)
p_mh- : significance level (assumption: underestimation of treatment effect)

As the results indicate that lifelong abstainers were at a higher risk of having died by June 2009, the fact that the Q_mh- statistic is significant for all values of Γ is not surprising, and indicate that the chance that the true treatment effect has been underestimated is quite low.

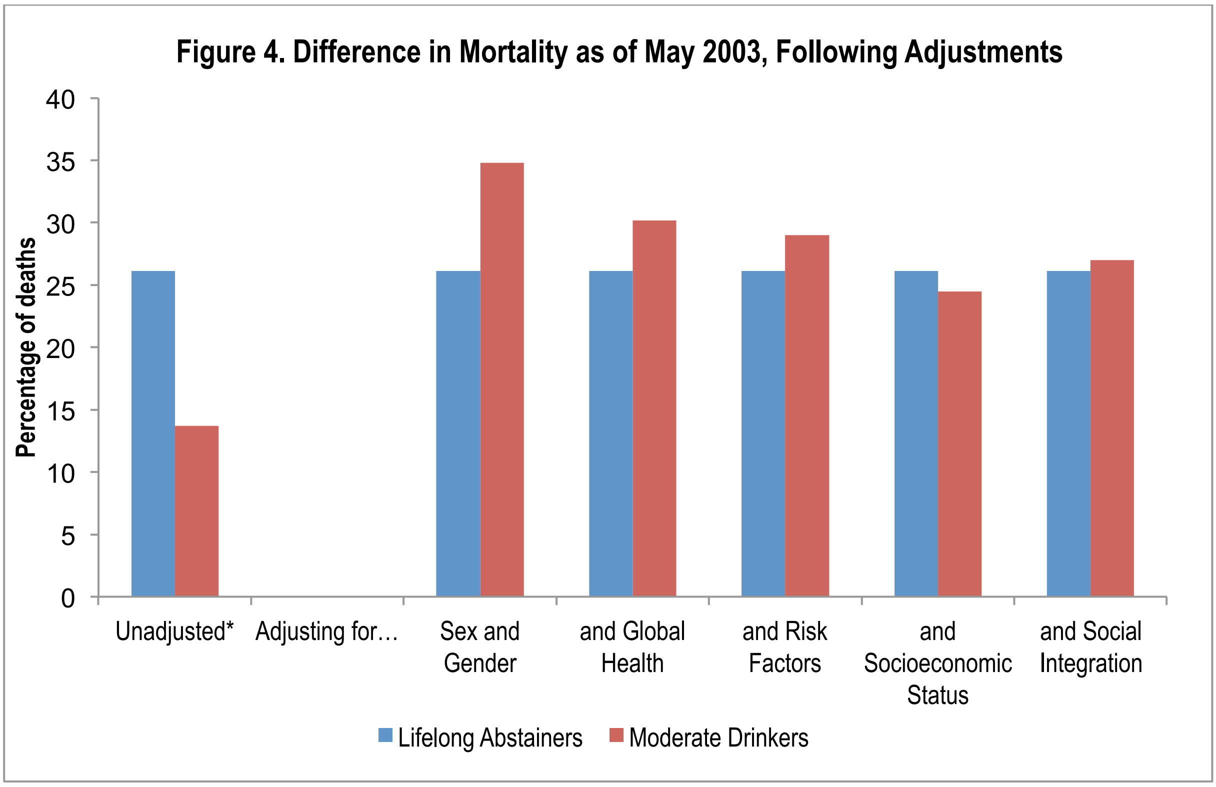
However, under the assumption that the treatment effect has been overestimated, the Q_mh+ statistic is no longer significant ($p<0.05$) at a value of $\Gamma = 1.25$. Thus, if an unobserved variable caused the odds ratio between groups to differ by 1.25, the confidence interval would include zero and the results would no longer be statistically significant.

4.4 Comparison to 2003

Melberg's original analysis was conducted using survival data generated through May 2003. The same analyses explored in sections 4.1 through 4.3 were repeated using mortality data generated through May 2003 as the endpoint. Naturally, the risk of dying before May 2003 was lower for both the moderate drinkers and the lifelong abstainers included in this sample, with 26.1% of lifelong abstainers having passed away by this point compared to only 13.7% of moderate drinkers.

4.4.1 NN Matching, with Replacement

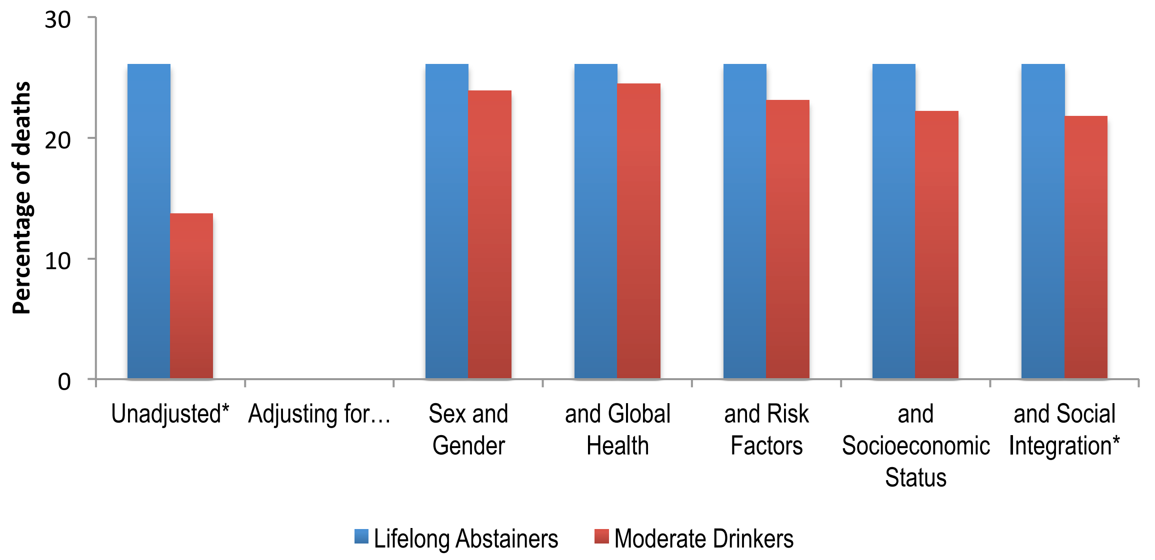
Following the same series of adjustments, the end result leaves lifelong abstainers 0.9% more likely to have been alive in May 2009 than their moderately drinking counterparts. However, this difference is not significant.



4.4.2 NN Matching Without Replacement.

Following the same procedure as outlined in section 4.2, the analysis was repeated without allowing for replacement. Following this series of adjustments, lifelong abstainers were 4.3% more likely to have died by May 2003 than their moderately drinking counterparts. While the differences in the mortality rates between the two groups ranged between 2% to 4% following each adjustment, it is only the final difference which was large enough to be considered significant. However, when comparing like with like, the mortality rate for moderate drinkers rose from 13.7% to 21.8%, which in turn diminishes the protective effect of moderate consumption from a difference in mortality rate of 12.4% to 4.3%.

Figure 5. Differences in Mortality as of May 2003, Following Adjustments (No Replacement)



As before, a sensitivity analysis was performed by computing Mantel-Haenszel bounds. The results are provided in Table 5

Table 5: Mantel-Haenszel (1959) bounds for The Risk of Dying Before May 2003				
Gamma (Γ)	Q_mh+	Q_mh-	p_mh+	p_mh-
1	1.98161	1.98161	**0.023761**	**0.023761**
1.05	1.56404	2.40095	0.058904	**8.18E-03**
1.1	1.1657	2.80079	0.121867	**2.55E-03**
1.15	0.785348	3.18349	0.216125	**7.28E-04**
1.2	0.421348	3.55058	0.33675	**1.92E-04**
1.25	0.072293	3.90336	0.471184	**4.70E-05**
1.3	0.145999	4.243	0.441961	**1.10E-05**
1.35	0.468593	4.57053	0.31968	**2.40E-06**
1.4	0.779525	4.88686	0.217835	**5.10E-07**
1.45	1.07966	5.19279	0.140147	**1.00E-07**
1.5	1.36976	5.48905	0.085381	**2.00E-08**

Gamma : odds of differential assignment due to unobserved factors
Q_mh+ : Mantel-Haenszel statistic (assumption: overestimation of treatment effect)

Q_mh- : Mantel-Haenszel statistic (assumption: underestimation of treatment effect)
p_mh+ : significance level (assumption: overestimation of treatment effect)
p_mh- : significance level (assumption: underestimation of treatment effect)

As before, the difference between the risk of dying before May 2003 for moderate drinkers and lifelong abstainers was significantly different. As expected, the corresponding Q_mh- statistic is significant for all values of Γ tested. Under the assumption that the treatment effect has been overestimated, the Q_mh+ statistic is no longer significant ($p<0.05$) at a value of $\Gamma = 1.05$. Thus, if an unobserved variable caused the odds ratio between groups to differ by the relatively minor factor of 1.05, the observed difference in risk would no longer be significant.

4.5 Regression Model (Basis of PS Estimation)

In these analyses, the propensity scores were estimated using a logistic regression model. All of the variables included in the model were selected based on the rationale provided in section 3.3. As noted previously, the literature was used to select characteristics of interest. Table 6 presents the model used.

Table 6: Logistic Regression Model Used to Estimate Probabilities for the Matching Procedure						
Log likelihood=-1465.6062				**Number of obs =2909** **LR chi 2(24) =504.29** **Prob >chi2 =0** **Pseudo R2= 0.1468**		
Nondrinkers	**Coef.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval**	
Age	**0.0497751**	**0.0042557**	**11.7**	**0.000**	**0.0414341**	**0.058116**
Gender	**-0.8928193**	**0.1144171**	**-7.8**	**0.000**	**-1.117073**	**-0.6685659**
Poor Health	**0.4206487**	**0.1090207**	**3.86**	**0.000**	**0.206972**	**0.6343255**
Low Activity Level	0.2289257	0.1713332	1.34	0.182	-0.1068813	0.5647327
Considers Self Overweight	0.1026717	0.1333989	0.77	0.442	-0.1587853	0.3641288
Prescription Drug Use 1984/85	0.1614118	0.1113124	1.45	0.147	-0.0567566	0.3795802
Prescription Drug Use 1991/92	-0.1185292	0.1108285	-1.07	0.285	-0.3357491	0.0986907
Restricted Diet 1984/85	-0.031632	0.1734399	-0.18	0.855	-0.371568	0.308304
Restricted Diet 1991/92	0.1400276	0.1510405	0.93	0.354	-0.1560064	0.4360616
Handicapped	**0.3796467**	**0.1471009**	**2.58**	**0.010**	**0.0913344**	**0.6679591**
Suffers from Chronic Illness	**-0.2917817**	**0.1310687**	**-2.23**	**0.026**	**-0.5486716**	**-0.0348917**
Reported High Blood Pressure	-0.0233473	0.1114442	-0.21	0.834	-0.2417739	0.1950792
Reported Heart Trouble	0.2309199	0.1815274	1.27	0.203	-0.1248673	0.586707
Smoked/Has Smoked	**-0.8656641**	**0.0960375**	**-9.01**	**0.000**	**-1.053894**	**-0.677434**
Treated for Depression	-0.192919	0.1188916	-1.62	0.105	-0.4259423	0.0401043
Left School Before 16	**0.3578659**	**0.1030227**	**3.47**	**0.001**	**0.155945**	**0.5597868**
Lower Status	**0.4043779**	**0.1284944**	**3.15**	**0.002**	**0.1525335**	**0.6562224**
Feelings of Loneliness	-0.1814832	0.1012746	-1.79	0.073	-0.3799778	0.0170115
Felt Unloved	**0.7676852**	**0.3024483**	**2.54**	**0.011**	**0.1748975**	**1.360473**
No Friends Nearby	0.2376077	0.1741938	1.36	0.173	-0.1038059	0.5790213
Felt Isolated-1991/92	-0.2541554	0.1499057	-1.7	0.090	-0.5479651	0.0396543
Felt Isolated -1984/85	0.0249589	0.1273075	0.2	0.845	-0.2245593	0.274477
Household Size 1984/85	**0.1107964**	**0.0457425**	**2.42**	**0.015**	**0.0211428**	**0.2004501**
Household Size 1991/92	0.0910005	0.0527221	1.73	0.084	-0.0123329	0.1943339
_cons	-3.346667	0.2656204	-12.6	0.000	-3.867274	-2.826061

** Variables in bold are significant (p<0.05)*

5 Discussion and Limitations

5.1 General Conclusions

The aim of this analysis was to determine if a propensity score analysis of moderate drinkers and lifelong abstainers would reveal the differences in all-cause mortality predicted by the J-curve hypothesis. These results do provide further support for the assertion that individuals classified as moderate drinkers and lifelong abstainers experience differences in mortality.

However, the results come with several caveats. First, as the bias reduction tables reveal, in this sample, individuals classified as moderate drinkers and lifelong abstainers differ in tangible ways. These differences are particularly apparent in the analyses where the option of replacement has been restricted, resulting in the failure of the matching algorithm to eliminate the bias associated with several variables, the most notable being age. It is worth noting that when replacement is allowed, the mortality outcomes in the two groups are no longer significantly different. Due to the weight the term "statistically significant" carries in the medical community, it is important to underscore the role that the method used, and the implementation of such a method, plays on the overall outcome.

A second goal of this analysis was to examine how an additional six years of mortality data has impacted the outcome. Obviously, fewer people had died by 2003 compared to 2009. Results of the analysis using 2003 mortality data when allowing for replacement closely mirror those from Melberg's earlier analysis – in both analyses, the matching algorithm shifted the difference in the risk of dying by roughly 1% in favor of the lifelong abstainers. By contrast, the 2009 analysis continued to reflect a higher risk of mortality among lifelong abstainers, though the difference in risk was not significant after adjustment. When examining the results from the analyses where replacement was restricted, the impact of additional time is even more interesting. While in both cases, matching without replacement confirmed that lifelong abstainers were at a higher risk of dying before the selected endpoint, the magnitude of this difference between the groups was quite different. Lifelong abstainers were 4.3% more likely to have passed away before moderate drinkers in 2003, but 9.8% more likely in 2009. This magnitude of difference is also mirrored in the analyses where replacement was allowed, suggesting that the risk of dying increased during this period. One possible explanation for this pattern is the difference in the age distribution between the two groups. Rerunning the analysis by excluding all individuals over 70 at HALS1 - in other words, excluding those who were most likely to have died by May 2009, regardless of lifestyle choice – results in a minor reduction in the mortality risk for the lifelong abstainers. However, the persistence of the higher risk indicates that the differences contributing to an increased risk of mortality is contingent on more than the difference in age.

5.2 Methods and Variables

This analysis attempted to adjust for several categories of covariates which have been associated with alcohol and mortality. Deciding which variables to include required a balancing act. While the bias reduction tables show that the individuals in the two groups differed significantly on a number of these variables, not all of the variables included had a statistically significant impact on mortality. Several were included because there is evidence to suggest that these are variables of interest, even if these relationships are not apparent in this dataset, and others were included because they were part of the original 2006 analysis. If the distribution of one of these variables, such as the presence or absence of treated depression, does not differ significantly among the two groups, and does not contribute significantly to the mortality risk, do the assumptions regarding the use of a logistic regression model make such variables suitable for use here? In this analysis, such variables have not been treated as superfluous due to the support in the literature for their inclusion.

However, it can also be argued that while there is support for the inclusion of variables which fall under the general areas of interest explored in this thesis, the particular variables selected are not a good representation of these qualities. It is entirely conceivable that treatment for depression, rather than serving as a risk factor, instead lowers the risk of an adverse outcome following treatment. Similarly, perhaps using more concrete measures of social inclusion, such as the number of visitors an individual had in the month prior to entering the HALS dataset would have offered better insight.

The decision to rely on less concrete measures was guided by the desire to try and exploit the measures available in the dataset which provided insight into more enduring characteristics – i.e. an individual who feels excluded from the community is likely to feel this way following events which have occurred over a longer time period, whereas the number of visitors an individual has can vary a great deal from week to week. Even so, the HALS dataset only includes measures from two points in time. Clearly, this poses challenges when attempting to assess more enduring traits. While some of these characteristics, such as gender or the age an individual left school can be considered permanent traits, there are several variables, such as whether or not an individual used prescription drugs, which merely provide information related to one particular instance in time. Naturally, in an ideal situation, repeated measurements would be available, so that it would be easier to draw conclusions about enduring traits. This is particularly true for variables which deal with things such as the history of a past illness. Individuals who had never smoked before 1991/1992 could have taken up the habit in 1993, but this information is simply impossible to trace.

One serious concern with this study is the risk of a misclassification error. As one of the objectives of this thesis was to revisit Melberg's 2006 analysis, individuals who self-identified as occasional drinkers were pooled with total abstainers in a lifelong abstainers category, similar to the category constructed by Melberg in his earlier analysis. While this inclusive approach creates a larger pool of abstainers for the analysis, it misclassifies special occasion drinkers as total abstainers. As a result, one can argue that the difference in mortality explored does not really show the true effect of lifelong abstention, but rather that of near abstention and abstention. Interestingly, running the analyses using only those individuals who identified as complete lifelong abstainers resulted in the risk of mortality between complete abstainers and moderate drinkers not differing statistically, even when the option of replacement was removed. However, as the group of these self-described total abstainers was very small (n=239, or 28.5% of lifelong abstainers and only 8.2% of the cases included in the analysis), reading too much into these results is not advisable.

Another issue present in this thesis is the possible presence of intermediary variables. The analyses used here have treated variables such as feelings of loneliness, heart problems, and high blood pressure as causal variables. In essence, this assumes that these factors influence alcohol consumption patterns and mortality. Yet suppose that some of these conditions are caused by alcohol consumption, or the lack thereof. Then, the inclusion of these variables leads to the true impact of an alcohol consumption pattern being underestimated. For example, if drinking is a social ritual, an individual who abstains could experience feelings of isolation stemming from abstention. In this scenario, abstention from alcohol would cause loneliness, and the combined effect would impact mortality. However, the sensitivity analyses performed indicate that the probability that the treatment effect has been underestimated is quite low; rather, there is a greater chance that the treatment effect has been overestimated. This suggests that, while intermediary variables may have been included in this analysis, their impact may be somewhat limited. This serves to underscore the complexity of the relationships between alcohol consumption, mortality, and the various factors which shape the decisions people take.

Was this study necessary? In the years following Melberg's 2006 findings, a number of studies have been conducted in this area, and the vast majority show that moderate alcohol consumption is associated with lower mortality. However, the debate continues. Many of the studies in this area are based on survival analysis, which is a method which comes with its own limitations. A Cox proportional hazards model assumes that the hazard ratio between the two groups remains constant over time. For example, Fuller's August 2011 study in *Demography*, which utilized Cox proportional hazards in the analysis, concluded that moderate alcohol consumption reduced mortality. However, the follow-up period of this study was merely 2-6 years, and it would be interesting to see how this assumption holds when using a much longer time horizon. Every method has its limitations, and while survival analysis offers many advantages, there is nothing to be lost by checking for the presence of this relationship through a variety of methods, particularly when the results shape policy guidelines and treatment.

5.3 Limitations

5.3.1 Limitations of the Method

Like all methods, propensity score matching has its limitations. First, in any dataset, there is always the risk that not all covariates have been observed. One hopes that the unobserved covariates correlate with observed covariates, but hope and certainty are not equivalent. While a sensitivity analysis can help to estimate the impact of these unobserved or excluded variables on the outcome, these analyses also offer estimates.

Secondly, propensity score matching is a data hungry method. In this analysis in particular, the ideal dataset would have been large enough for good matches to be available for all members of the treatment group. As matters stand, the only way to find good matches for all individuals in the lifelong abstainers group is to accept the increased variance which comes when matching without replacement.

Finally, propensity score matching is dependent on a series of subjective decisions, in terms of the data selected, the variables included, and the transformations performed. Often, these decisions are made somewhat arbitrarily. While the variables included in this analysis were selected based on insight from the literature, it is impossible to escape the fact that there has been a certain amount of subjective decision-making. The ways in which the analysis utilized here deviates from Melberg's initial analysis only underscores this fact. The same dataset, approached slightly differently, yields slightly different outcomes. Undoubtedly, if this analysis were repeated in another 5 years time, the approach would vary slightly again.

5.3.2 Limitations of the Data

While the HALS dataset is quite extensive and offers several advantages, it also offers some distinct disadvantages, particularly within the framework of this particular analysis. One issue is the problem of missing data, both in terms of the cases lost between HALS1 and 2, and in terms of certain variables having a high percentage of missing data. Some of the variables missing a great deal of information are ones which encompass information of particular interest when assessing mortality risk. The example of BMI springs to mind, as this is a variable of interest where values are missing for 10% of the population. Likewise, income is another piece of information which has been linked to differences in mortality risk where data is missing for nearly 20% of the sample population.

A second limitation is also linked to missing information, but where time is the main cause. While s higher percentage of the sample had passed away by June 2009 compared to 2003, the majority of the sample is still alive. Ideally, this is an analysis which should be conducted when the entire population has passed away. As of now, we have chosen a somewhat arbitrary endpoint. Furthermore, as a majority of the sample is still alive, it is impossible to utilize all the information the HALS datasets provide. For example, the HALS dataset does allow for the possibility of more specialized measures of mortality risk, such as the risk of dying of cardiovascular disease, or the risk of dying by a given age. However, as a large proportion of the sample is still alive, these analyses are not entirely relevant. For example, less than 5% of the individuals who have passed away so far had coronary heart disease listed as the primary or underlying cause of death. While it was originally hoped that a comparison between the mortality risk in general and heart disease in particular would be possible, there is simply not enough data for use in a propensity score based analysis.

This is not a limitation of this analysis alone, but is also a weakness within the area of survival analysis. As this analysis has shown, 6 additional years does make a difference in the results. While it appears that the additional time only serves to support the trends seen in the earlier analyses, there is an argument for allowing for the passage of more time before setting these conclusions in stone.

6 References

Becker, S.O. and Caliendo, M. (2007) “Sensitivity analysis for average treatment effect.” *Stata Journal* **7**(1): 71–83.

Benjamins, M. R., Hummer, R. A., Eberstein, I. W., & Nam, C. B. (2004) ”Self-reported health and adult mortality: An analysis of cause-specific mortality.” *Social Science and Medicine* 59, 1297–1306.

Boston University Medical Center. "What should we advise about alcohol consumption? A debate amongst scientists." *ScienceDaily*, 18 Jan. 2011. Web. 05 Feb. 2011.

Caliendo, M. & Kopeinig, S. (2008) “Some Practical Guidance For The Implementation of Propensity Score Matching.” *Journal of Economic Surveys* 22: 31–72. doi: 10.1111/j.1467-6419.2007.00527.x

Corrao, G., Rubbiati, L., Bagnardi, V., Zambon, A., & Poikolainen, K. (2000) “Alcohol and Coronary Heart Disease: A Meta Analysis.” *Addiction* 95 (10):1505-1523

Cox, B.D. (1988) “Health and Lifestyle Survey, 1984-1985.” [computer file]. Colchester, Essex: UK Data Archive [distributor], October 1988. SN: 2218, http://dx.doi.org/10.5255/UKDA-SN-2218-1.

Cox, B.D. (1995) “Health and Lifestyle Survey: Seven-Year Follow-Up, 1991-1992.” [computer file]. Colchester, Essex: UK Data Archive [distributor], January 1995. SN: 3279.

Cox, B.D. (2009) “Health and Lifestyle Survey Deaths and Cancer Data, June 2009” [computer file]. Colchester, Essex: UK Data Archive [distributor], November 2009. SN: 6339, http://dx.doi.org/10.5255/UKDA-SN-6339-1.

D'Agostino Jr, R.B. (1998)”Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group.” *Statistics in Medicine* (17). 2265–2281

Del Boca, F. K. & Darkes, J. (2003) “The validity of self-reports of alcohol consumption: state of the science and challenges for research.” *Addiction* 98:1-12. doi: 10.1046/j.1359-6357.2003.00586.x

Di Castelnuovo A., Costanzo S., Bagnardi V., Donati M.B., Iacoviello L., de Gaetano G., (2006) “Alcohol dosing and total mortality in men and women: an updated meta-analysis of 34 prospective studies.” *Archives of Internal Medicine* (166):2437-2445.

Ellison, R.C. (2002), ”Balancing the risks and benefits of moderate drinking.” *Annuals of the New York Academy of Sciences* 957:1-6

Feinstein, J.S.(1993), “The Relationship between Socioeconomic Status and Health: A Review of the Literature” *The Milbank Quarterly* 71(2): 279-322

Fillmore, K. M., Kerr, W. C., Stockwell, T., Chikritzhs, T., & Bostrom, A. (2006) “Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies.” *Addiction Research and Theory* 14, 101–132.

Fuller, T.D., (2011). “Moderate alcohol consumption and the risk of mortality.” *Demography* 48(3): 1105-1125.

Goetghebeur E, & Pocock S.J., (1995) “Detection and estimation of J-shaped risk-response relationships.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 158(1): 107–121

Heckman, J. (1979) "Sample selection bias as a specification error". *Econometrica* **47** (1): 153–61. doi:10.2307/1912352. JSTOR 1912352

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998) "Characterizing selection bias using experimental data." *Econometrica* 66(5): 1017-1098.

Holland, P. (1986) “Statistics and causal inference (with discussion).” *Journal of the American Statistical Association* 81, 945-970.

Holt-Lunstad J, Smith T.B, & Layton J.B., (2010) “Social Relationships and Mortality risk: a meta-analytic review.” *PLoS Med* 7(7):1-20. doi:10.1371/journal.pmed.1000316

Idler, E. L., & Benyamani, Y. (1997) ”Self-rated health and mortality: A review of twenty-seven community studies.” *Journal of Health and Social Behavior*, 38, 21–37.

Israel, Y., Hollander, O., Sanchez-Craig, M., Booker, S., Miller, V., Gingrich, R. & Rankin, J. G. (1996) “Screening for problem drinking and counseling by the primary care physician-nurse team.” *Alcoholism: Clinical and Experimental Research* 20: 1443–1450.

Leuven, E. and Sianesi, B. (2003) "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." http://ideas.repec.org/c/boc/bocode/s432001.html. Version 4.0.4 10 November 2010

Klatsky A.L. (2002) “Alcohol and cardiovascular diseases: a historical overview.” *Annuals of the New York Academy of Sciences* 957: 7–15

Magnus P., Bakke, E., Hoff, D.A., Høiseth, G., Graff-Iversen S., Knudsen, G.P., Myhre R., Normann P.T., Næss O., Tambs K., Thelle D.S., Mørland J. (2011) “Controlling for high-density lipoprotein cholesterol does not affect the magnitude of the relationship between alcohol and coronary heart disease.” *Circulation* 124:2296–2302.

Melberg, Hans Olav (2006). “Does moderate alcohol intake reduce mortality?” In Jon Elster; Olav Gjelsvik; Aanund Hylland; Karl-Ove Moene & Hans Olav Melberg (ed.), *Understanding Choice, Explaining Behaviour: Essays in Honour of Ole-Jørgen Skog.* Unipub forlag. ISBN 82-7477-237-7. 12. 191 - 210

Midanik, L. T. (1988) “Validity of self-reported alcohol use: a literature review and assessment.” *British Journal of Addiction*, 83: 1019–1029. doi: 10.1111/j.1360-0443.1988.tb00526.

NHS (2011) “Alcohol Units” [online] Available at; http://www.nhs.uk/Livewell/alcohol/Pages/alcohol-units.aspx last revised 26/4/2011, last accessed September 20 2011

Ponz de Leon, M., (2010) “What should we advise about alcohol consumption?” *Internal and Emergency Medicine* 6(1);87 DOI:10.1007/s11739-010-0487-1

Rosenbaum, P. R. (1999) “Choice as an alternative to control in observational studies.” *Statistical Science* 14:259–278.

Rosenbaum, P. R. & Rubin, D. B. (1983) “The central role of the propensity score in observational studies for causal effects.” *Biometrika* 70: 41-55

Standridge, J. B., Zylstra, R. G., & Adams, S. M. (2004) “Alcohol consumption: An overview of benefits and risks.” *Southern Medical Journal* 97, 664–672

Stanistreet D., Scott-Samuel, A., & Bellis M.A. (1999) Income inequality and mortality in England.” *Journal of Public Health Medicine* 21: 205–207.

Temple, M., and Fillmore, K.M. (1985) “The variability of drinking patterns and problems among men, age 16–31: A longitudinal study.” *International Journal of the Addictions* 20:1595–1620.

Todd, Petra E. (2008). "Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated," *Handbook of Development Economics,* T. Paul Schultz & John A. Strauss (ed.) , Elsevier, chapter 60, pages 3847-3894

Winship, C. & Morgan, S.L., (1999) “The estimation of causal effects from observational data.” *Annual Review of Sociology* 25: 659–707

7 Appendix

7.1 Overview of Covariates and HALS variables

Overview of Covariates and HALS Variables					
Variable used in Analysis	**Description**	**Variable type (binary/continuous)**	**HALS1 Variable name**	**Corresponding HALS2 Variable (if used)**	**Description of HALS Variable(s)**
		General			
Age	age at time of survey	Continuous	agyrs		Age at time of survey
Gender	gender at HALS1	Binary (1=male)	sex	-	Gender at HALS1
		Self Rated Global Health			
Poor Health	Respondent assessment of his/her own health as poor in relation to others of the same age and gender at HALS1 or 2	Binary (1=Poor Health)	ownh	ownh2	Respondent's opinion of own health in relation to others of the same age.
		Risk Factors			
Low Activity Level	Respondent's perception of their own activity level compared to others of the same age and gender?	Binary(1=Less active than others of the same age and gender)	compact		Respondent's perception of their own activity level compared to others of the same age and gender
Considers Self Overweight	Respondent considered himself/herself overweight at HALS1 & 2	Binary (1=considers self overweight)	asswt	asswt200	Respondent considers himself/herself overweight.
			loswtdr		Has been advised to lose weight by a medical professional
Prescription Drug Use . 1984/85	Does respondent have a prescription other than contraceptives	Binary (1=has prescription)	drug		Does respondent have a prescription other than contraceptives
Prescription Drug Use- 1991/92	Does respondent have a prescription other than contraceptives	Binary (1=has prescription)		drug2	
Restricted Diet (1984/85)	Is responded on diet restrictions for health reasons	Binary (1=on diet)	diet		Is the responded on a restricted diet for health reasons?
Restricted Diet (1991/92)	Is responded on diet restrictions for health reasons	Binary (1=on diet)		diet200	

Table continued					
Variable used in Analysis	**Description**	**Variable type (binary/continuous)**	**HALS1 Variable name**	**Corresponding HALS2 Variable (if used)**	**Description of HALS Variable(s)**
Handicapped	Individual suffers from a handicap they consider limiting at HALS1 &HALS2	Binary (1=has handicap)	handcp	hcap200	Individual suffers from a handicap they consider limiting
Suffers from Chronic Illness	Chronic illness present at both HALS1 and HALS2	Binary (1=Chronic Illness present)	dis	dis200	Presence or absence of a long-term, chronic illness, including conditions such as heart trouble, kidney trouble, mental health issues, etc.
Reported High Blood Pressure	Reported history of high blood pressure	Binary (1=has had condition)	pastds14	pastds33	Has the respondent ever suffered from high blood pressure?
Reported Heart Trouble	reported history of heart trouble	Binary (1=has had condition)	pastds9	HEART1	Has the respondent ever had heart trouble?
Smokes/Has Smoked	Reported history of smoking	Binary (1=smoked)		smever	Has the respondent ever smoked, as more than an experiment?
Treated For Depression	Has been treated for depression	Binary (1=has had condition)	pastds12	PASTDS31	Have the respondent ever been treated for depression?
Socioeconomic Status					
Left School before 16	Identification of Individuals who left school before turning 16 (the age where schooling ceases to be mandatory in the UK)	Binary(1=left school before 16)	aglsch		Age respondent left school.
Lower Status	Social status based on RGSC. Individuals who fell into the two lowest social classes (semi skilled or unskilled labor) were classified as having lower status	Binary (1=Falls into the lower social classes at HALS1 &2)	shtseg1	shtseg2	Social status based on RGSC and related factors

Table continued					
Variable used in Analysis	**Description**	**Variable type (binary/continuous)**	**HALS1 Variable name**	**Corresponding HALS2 Variable (if used)**	**Description of HALS Variable(s)**
Social Integration					
Feelings of Loneliness	Individual reported feeling lonely often or constantly at HALS1 or 2	Binary (1=Felt Lonely)	MSYMPT8	MSYMTZ08	How often does the respondent feel lonely?
Felt Unloved	Individuals who reported that they had no people in their lives who made them feel loved at HALS1 or HALS2	Binary (1=felt unloved)	pssi1	PSSi201	Degree to which the respondent agrees with the statement "There are people who make me feel loved"
No Friends Nearby	Respondent does not have any friends in the area	Binary (1=No friends in area)		FRHERE2	Presence of friends in the respondent's community/life
Felt Isolated 1991/92	Respondent reported feeling isolated from the community	Binary (1=Felt Isolated)		parther2	Degree to which respondent feels part of the community
Felt Isolated 1984/85	Respondent reported feeling isolated from the community	Binary (1=Felt Isolated)	parthere		
Household Size - 1984/85	Number of people respondent lived with	Continuous	hou		Number of people respondent lives with
Household Size - 1991/92	Number of people respondent lives with	Continuous		HOU2	